Strong Converse Theorems for Degraded Broadcast Channels with Feedback

Yasutada Oohama
University of Electro-Communications, Tokyo, Japan
Email: oohama@uec.ac.jp

***Abstract*—We consider the discrete memoryless degraded broadcast channels with feedback. We prove that the error probability of decoding tends to one exponentially for rates outside the capacity region and derive an explicit lower bound of this exponent function. We shall demonstrate that the information spectrum approach is quite useful for investigating this problem.**

I. DBC WITH FEEDBACK

Let $\mathcal{X}, \mathcal{Y}$, and $\mathcal{Z}$ be finite sets. The broadcast channel we study in this paper is defined by a discrete memoryless channel specified with the following stochastic matrix:

$$W \stackrel{\triangle}{=} \{W(y,z|x)\}_{(x,y,z)\in\mathcal{X}\times\mathcal{Y}\times\mathcal{Z}}. \tag{1}$$

Here $\mathcal{X}$ is a set of channel input and $\mathcal{Y}$, and $\mathcal{Z}$ are sets of two channel outputs. We assume that those are finite sets. Let X^n be a random variable taking values in $\mathcal{X}^n$. We write an element of $\mathcal{X}^n$ as $x^n = x_1 x_2 \cdots x_n$. Suppose that X^n has a probability distribution on $\mathcal{X}^n$ denoted by $p_{X^n} = \{p_{X^n}(x^n)\}_{x^n\in\mathcal{X}^n}$. Similar notations are adopted for other random variables. Let $Y^n \in \mathcal{Y}^n$ and $Z^n \in \mathcal{Y}^n$ be random variables obtained as the channel output by connecting X^n to the input of channel. We write a conditional distribution of (Y^n, Z^n) on given X^n as

$$W^n = \{W^n(y^n, z^n|x^n)\}_{(x^n,y^n,z^n)\in\mathcal{X}^n\times\mathcal{Y}^n\times\mathcal{Z}^n}.$$

Since the channel is memoryless, we have

$$W^n(y^n, z^n|x^n) = \prod_{t=1}^{n} W(y_t, z_t|x_t). \tag{2}$$

In this paper we deal with the case where the components $W(z,y|x)$ of W satisfy the following conditions:

$$W(y,z|x) = W_1(y|x)W_2(z|y). \tag{3}$$

In this case we say that the broadcast channel W is *degraded*. The degraded broadcast channel (DBC) is specified by (W_1, W_2). Let K_n and L_n be uniformly distributed random variables taking values in message sets $\mathcal{K}_n$ and $\mathcal{L}_n$, respectively. The random variable K_n is a message sent to the receiver 1. The random variable L_n is a message sent to the receiver 2. In this paper we consider the case where we have feedback links from the receivers 1 and 2 to the sender. Transmission of the message pair (K_n, L_n) via the DBC with feedback is shown in Fig. 1. A feedback encoder denoted by $\tilde{\varphi}^n = \{\tilde{\varphi}_t\}_{t=1}^n$ consists of n encoder functions $\tilde{\varphi}_t$, $t = 1, 2, \cdots, L$, where for each $t = 1, 2, \cdots, n$,

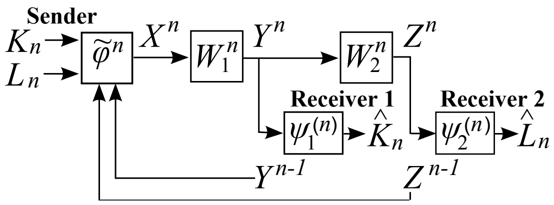

Fig. 1. Transmission of the message pair (K_n, L_n) via the DBC with feedback.

$$\tilde{\varphi}_t : \mathcal{K}_n \times \mathcal{L}_n \times \mathcal{Y}^{t-1} \times \mathcal{Z}^{t-1} \to \mathcal{X}_t$$

is a stochastic matrix. For a given message pair $(k,l) \in \mathcal{K}_n \times \mathcal{L}_n$ and given feedback signals $y^{n-1} \in \mathcal{Y}^n$ form the receiver 1 and $z^{n-1} \in \mathcal{Z}^n$ from the receiver 2, conditional provability of $x^n \in \mathcal{X}^n$ by $\tilde{\varphi}^n$ is

$$\tilde{\varphi}^n(x^n|k,l,y^{n-1},z^{n-1}) = \prod_{t=1}^{n} \tilde{\varphi}_t(x_t|k,l,y^{t-1},z^{t-1}).$$

The t-th transmission in the DBC with feedback is shown in Fig. 2. The joint probability mass function on $\mathcal{K}_n \times \mathcal{L}_n \times \mathcal{X}^n \times \mathcal{Y}^n \times \mathcal{Z}^n$ is given by

$$\begin{aligned}
&\mathrm{Pr}\{(K_n, L_n, X^n, Y^n, Z^n) = (k,l,x^n,y^n,z^n)\} \\
&= \frac{1}{|\mathcal{K}_n||\mathcal{L}_n|} \prod_{t=1}^{n} \{\tilde{\varphi}_t(x_t|k,l,y^{t-1},z^{t-1}) \\
&\qquad \times W_1(y_t|x_t) W_2(z_t|y_t)\},
\end{aligned}$$

where $|\mathcal{K}_n|$ is a cardinality of the set $\mathcal{K}_n$. We set

$$\begin{aligned}
&\tilde{p}_{K_n L_n X^n Y^n Z^n}(k,l,x^n,y^n,z^n) \\
&\stackrel{\triangle}{=} \mathrm{Pr}\{(K_n, L_n, X^n, Y^n, Z^n) = (k,l,x^n,y^n,z^n)\}.
\end{aligned}$$

By an elementary calculation we can show that for each $(l, x^n, y^n, z^n) \in \mathcal{L}_n \times \mathcal{X}^n \times \mathcal{Y}^n \times \mathcal{Z}^n$, the probability $\tilde{p}_{L^n X^n Y^n Z^n}(l, x^n, y^n, z^n)$ is given by

$$\begin{aligned}
&\tilde{p}_{L^n X^n Y^n Z^n}(l,x^n,y^n,z^n) = \tilde{p}_{L^n}(l) \\
&\times \prod_{t=1}^{n} \{\tilde{p}_{X_t|L_n X^{t-1} Y^{t-1} Z^{t-1}}(x_t|l,x^{t-1},y^{t-1},z^{t-1}) \\
&\qquad \times W_1(y_t|x_t) W_2(z_t|y_t)\}.
\end{aligned}$$

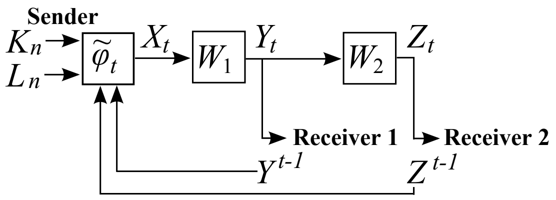

Fig. 2. The t-th transmission in the DBC with feedback.

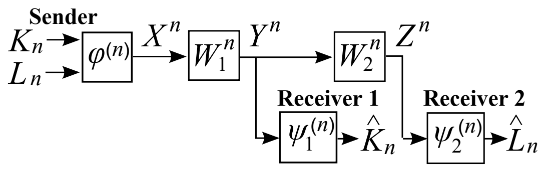

Fig. 3. Transmission of messages via the degraded BC.

The decoding functions at the receiver 1 and the receiver 2, respectively, are denoted by $\psi_1^{(n)}$ and $\psi_2^{(n)}$. Those functions are formally defined by $\psi_1^{(n)} : \mathcal{Y}^n \to \mathcal{K}_n, \psi_2^{(n)} : \mathcal{Z}^n \to \mathcal{L}_n$. The average error probability of decoding on the receivers 1 and 2 is defined by

$$\begin{aligned} &\mathrm{P}_{\mathrm{e,FB}}^{(n)} = \mathrm{P}_{\mathrm{e,FB}}^{(n)}(\varphi^{(n)}, \psi_1^{(n)}, \psi_2^{(n)}) \\ &\stackrel{\triangle}{=} \Pr\{\psi_1^{(n)}(Y^n) \neq K_n \text{ or } \psi_2^{(n)}(Z^n) \neq L_n\} \end{aligned}$$

For $k \in \mathcal{K}_n$ and $l \in \mathcal{L}_n$, set $\mathcal{D}_1(k) \stackrel{\triangle}{=} \{y^n : \psi_1^{(n)}(y^n) = k\}$, $\mathcal{D}_2(l) \stackrel{\triangle}{=} \{z^n : \psi_2^{(n)}(z^n) = l\}$. The families of sets $\{\mathcal{D}_1(k)\}_{k\in\mathcal{K}_n}$ and $\{\mathcal{D}_2(l)\}_{l\in\mathcal{L}_n}$ are called the decoding regions. Using the decoding region, $\mathrm{P}_{\mathrm{e,FB}}^{(n)}$ can be written as

$$\begin{aligned} &\mathrm{P}_{\mathrm{e,FB}}^{(n)} = \mathrm{P}_{\mathrm{e,FB}}^{(n)}(\varphi^n, \psi_1^{(n)}, \psi_2^{(n)}) \\ &= \frac{1}{|\mathcal{K}_n||\mathcal{L}_n|} \sum_{(k,l)\in\mathcal{K}_n\times\mathcal{L}_n} \sum_{\substack{(x^n,y^n,z^n)\in\mathcal{X}^n\times\mathcal{Y}^n\times\mathcal{Z}^n: \\ y^n\in\mathcal{D}_1(k) \text{ or } z^n\in\mathcal{D}_2(l)}} \\ &\quad\times\tilde{\varphi}^n(x^n|k,l,y^{n-1},z^{n-1})W_1^n(y^n|x^n)W_2^n(z^n|y^n). \end{aligned}$$

The average correct probability of decoding is defined by

$$\mathrm{P}_{\mathrm{c,FB}}^{(n)} = \mathrm{P}_{\mathrm{c,FB}}^{(n)}(\tilde{\varphi}^n, \psi_1^{(n)}, \psi_2^{(n)}) = 1 - \mathrm{P}_{\mathrm{e,FB}}^{(n)}(\tilde{\varphi}^n, \psi_1^{(n)}, \psi_2^{(n)}).$$

On the other hand, transmission of messages via the DBC without feedback is shown in Fig. 3. In this figure, $\varphi^{(n)}$ is a stochastic matrix given by

$$\varphi^{(n)} = \{\varphi^{(n)}(x^n|k,l)\}_{(k,l,x^n)\in\mathcal{K}_n\times\mathcal{L}_n\times\mathcal{X}^n},$$

where $\varphi^{(n)}(x^n|k,l)$ is a conditional probability of $x^n \in \mathcal{X}^n$ given message pair $(k,l) \in \mathcal{K}_n \times \mathcal{L}_n$. Let the average error probability of decoding in the case without feedback be denoted by $\mathrm{P}_{\mathrm{e}}^{(n)}$. This quantity has the following form

$$\begin{aligned} \mathrm{P}_{\mathrm{e}}^{(n)} = \frac{1}{|\mathcal{K}_n||\mathcal{L}_n|} &\sum_{(k,l)\in\mathcal{K}_n\times\mathcal{L}_n} \sum_{\substack{(x^n,y^n,z^n)\in\mathcal{X}^n\times\mathcal{Y}^n\times\mathcal{Z}^n: \\ y^n\in\mathcal{D}_1(k) \text{ or } z^n\in\mathcal{D}_2(l)}} \\ &\times\varphi^{(n)}(x^n|k,l)W_1^n(y^n|x^n)W_2^n(z^n|y^n). \end{aligned}$$

The average correct probability of decoding is defined by

$$\mathrm{P}_{\mathrm{c}}^{(n)} = \mathrm{P}_{\mathrm{c}}^{(n)}(\varphi^{(n)}, \psi_1^{(n)}, \psi_2^{(n)}) \stackrel{\triangle}{=} 1 - \mathrm{P}_{\mathrm{e}}^{(n)}(\varphi^{(n)}, \psi_1^{(n)}, \psi_2^{(n)}).$$

For $\varepsilon \in (0,1)$, a pair (R_1, R_2) is ε-*achievable* if there exists a sequence of triples $\{(\tilde{\varphi}^n, \psi_1^{(n)}, \psi_2^{(n)})\}_{n=1}^{\infty}$ such that

$$\begin{aligned} &\mathrm{P}_{\mathrm{e,FB}}^{(n)}(\tilde{\varphi}^n, \psi_1^{(n)}, \psi_2^{(n)}) \leq \varepsilon, \\ &\liminf_{n\to\infty} \frac{1}{n} \log|\mathcal{K}_n| \geq R_1, \liminf_{n\to\infty} \frac{1}{n} \log|\mathcal{L}_n| \geq R_2. \end{aligned}$$

The set that consists of all ε-achievable rate pair is denoted by $\mathcal{C}_{\mathrm{DBC,FB}}(\varepsilon|W_1, W_2)$. Furthermore, set

$$\mathcal{C}_{\mathrm{DBC,FB}}(W_1, W_2) = \bigcap_{\epsilon\in(0,1)} \mathcal{C}_{\mathrm{DBC,FB}}(\varepsilon|W_1, W_2).$$

We define the capacity region $\mathcal{C}_{\mathrm{DBC}}(\varepsilon|W_1, W_2)$ in the case without feedback in a manner quite similar to the definition of $\mathcal{C}_{\mathrm{DBC,FB}}(\varepsilon|W_1, W_2)$. We define the capacity region $\mathcal{C}_{\mathrm{DBC}}(W_1, W_2)$ of the DBC without feedback in a manner quite similar to the definition of $\mathcal{C}_{\mathrm{DBC,FB}}(W_1, W_2)$.

To describe $\mathcal{C}_{\mathrm{DBC}}(W_1, W_2)$, we introduce an auxiliary random variable U taking values in a finite set $\mathcal{U}$. We assume that the joint distribution of (U, X, Y, Z) is

$$\begin{aligned} &p_{UXYZ}(u, x, y, z) \\ &= p_U(u)p_{X|U}(x|u)W_1(y|x)W_2(z|y). \end{aligned}$$

The above condition is equivalent to $U \leftrightarrow X \leftrightarrow Y \leftrightarrow Z$. Define the set of probability distribution $p = p_{UXYZ}$ of $(U, X, Y, Z) \in \mathcal{U} \times\mathcal{X} \times\mathcal{Y} \times\mathcal{Z}$ by

$$\begin{aligned} \mathcal{P}(W_1, W_2) \stackrel{\triangle}{=} \{&p : |\mathcal{U}| \leq |\mathcal{X}| + 1, \\ &p_{Y|X} = W_1, p_{Z|Y} = W_2, U \leftrightarrow X \leftrightarrow Y \leftrightarrow Z\}. \end{aligned}$$

Set

$$\begin{aligned} \mathcal{C}(p) &\stackrel{\triangle}{=} \{(R_1, R_2) : R_1, R_2 \geq 0\,, \\ &\qquad R_1 \leq I_p(X; Y|U), R_2 \leq I_p(U; Z)\}. \\ \mathcal{C}(W_1, W_2) &= \bigcup_{p\in\mathcal{P}(W_1,W_2)} \mathcal{C}(p). \end{aligned}$$

The broadcast channel was posed investigated by Cover [1]. Previous results on the capacity region for the DBC are given by the following theorem.

Theorem 1 ([2]-[5]): For each fixed $\varepsilon \in (0,1)$ and any DBC (W_1, W_2), we have

$$\begin{aligned} &\mathcal{C}_{\mathrm{DBC}}(\varepsilon|W_1, W_2) = \mathcal{C}_{\mathrm{DBC}}(W_1, W_2) \\ &= \mathcal{C}(W_1, W_2). \end{aligned}$$

A previous result on $\mathcal{C}_{\mathrm{DBC,FB}}(W_1, W_2)$ is given by the following theorem stating that the feedback can not increase the capacity region for the DBC.

Theorem 2 (El Gamal [7]): For any DBC $(W_1, W_2\ )$, we have

$$\begin{aligned} \mathcal{C}_{\mathrm{DBC,FB}}(W_1, W_2) &= \mathcal{C}_{\mathrm{DBC}}(W_1, W_2) \\ &= \mathcal{C}(W_1, W_2). \end{aligned}$$

In general broadcast channels the feedback can increase the capacity region. Previous works on the coding problem for broad cast channels with feedback are summarized in [8].

To examine an asymptotic behavior of $\mathrm{P}_{\mathrm{c,FB}}^{(n)}$ for rate pairs outside the capacity region we define the following quantity.

$$
\begin{aligned}
&G_{\mathrm{FB}}^{(n)}(R_1,R_2|W_1,W_2)\\
&\triangleq \min_{\substack{(\tilde{\varphi}^n,\psi_1^{(n)},\psi_2^{(n)}):\\ (1/n)\log|\mathcal{K}_n|\geq R_1,\\ (1/n)\log|\mathcal{L}_n|\geq R_2}} \left(-\frac{1}{n}\right)\log \mathrm{P}_{\mathrm{c,FB}}^{(n)}(\tilde{\varphi}^n,\psi_1^{(n)},\psi_2^{(n)}),\\
&G_{\mathrm{FB}}(R_1,R_2|W_1,W_2)=\lim_{n\to\infty}G_{\mathrm{FB}}^{(n)}(R_1,R_2|W_1,W_2).
\end{aligned}
$$

The quantity $G_{\mathrm{FB}}(R_1,R_2|W_1,W_2)$ is the optimal exponent function for the correct probability of decoding at rate pairs outside the capacity region. In the case without feedback we define the optimal exponent function $G(R_1,R_2|W_1,W_2)$ for the correct probability of decoding for rate pairs outside the capacity region in a manner quite similar to the definition of $G_{\mathrm{FB}}(R_1,R_2|W_1,W_2)$.

Define

$$
\begin{aligned}
&\omega_q^{(\mu)}(x,y,z|u)\\
&\triangleq \mu\log\frac{q_{Y|X}(y|x)}{q_{Y|U}(y|u)}+\log\frac{q_{Z|U}(z|u)}{q_Z(z)},\\
&\Lambda_q^{(\mu,\lambda)}(XYZ|U)\\
&\triangleq \sum_{(u,x,y,z)\in\mathcal{U}\times\mathcal{X}\times\mathcal{Y}\times\mathcal{Z}} q_{UX}(u,x)q_{Y|X}(y|x)q_{Z|Y}(z|y)\\
&\qquad\times\exp\left\{\lambda\omega_q^{(\mu)}(x,y,z|u)\right\},\\
&\Omega_q^{(\mu,\lambda)}(XYZ|U)\triangleq\log\Lambda_q^{(\mu,\lambda)}(XYZ|U),\\
&\Omega^{(\mu,\lambda)}(W_1,W_2)\triangleq\max_{q\in\mathcal{P}(W_1,W_2)}\Omega_q^{(\mu,\lambda)}(XYZ|U),\\
&F^{(\mu,\lambda)}(\mu R_1+R_2|W_1,W_2)\\
&\triangleq\frac{\lambda(\mu R_1+R_2)-\Omega^{(\mu,\lambda)}(W_1,W_2)}{1+2\lambda+\lambda\mu},\\
&F(R_1,R_2|W_1,W_2)\\
&\triangleq\sup_{\mu,\lambda>0}\frac{\lambda(\mu R_1+R_2)-\Omega^{(\mu,\lambda)}(W_1,W_2)}{1+2\lambda+\lambda\mu}.
\end{aligned}
$$

We can show that the above functions and sets satisfy the following property.

Property 1:

a) For each $q\in\mathcal{P}(W_1,W_2)$, $\Omega_q^{(\mu,\lambda)}(XYZ|U)$ is a monotone increasing and convex function of $\lambda>0$.

b) For every $q\in\mathcal{P}(W_1,W_2)$, we have
$$\lim_{\lambda\to+0}\frac{\Omega_q^{(\mu,\lambda)}(XYZ|U)}{\lambda}=\mu I_q(X;Y|U)+I_q(U;Z).$$

c) If $(R_1,R_2)\notin\mathcal{C}(W_1,W_2)$, then we have $F(R_1,R_2|W_1,W_2)>0$.

The author [6] obtained the following.

Theorem 3: For any DBC (W_1,W_2), we have
$$G(R_1,R_2|W_1,W_2)\geq F(R_1,R_2|W_1,W_2). \tag{4}$$

It follows from Theorem 3 and Property 1 part c) that if (R_1,R_2) is outside the capacity region, then the error probability of decoding goes to one exponentially and its exponent is not below $F(R_1,R_2|W_1,W_2)$.

Our result in the case of feedback is the following.

Theorem 4: For any DBC (W_1,W_2), we have
$$G_{\mathrm{FB}}(R_1,R_2|W_1,W_2)\geq F(R_1,R_2|W_1,W_2). \tag{5}$$

It is interesting that the exponent function $F(R_1,R_2|W_1,W_2)$ also serves as a lower bound of the optimal exponent function $G_{\mathrm{FB}}(R_1,R_2|W_1,W_2)$ in the case of feedback. This result strongly suggests a possibility that the feedback can not improve the optimal exponent function for the probability of correct decoding at the rate pairs outside the capacity region.

From this theorem we immediately follows from the following corollary.

Corollary 1: For each fixed $\varepsilon\in(0,1)$, and any DBC (W_1,W_2), we have
$$
\begin{aligned}
&\mathcal{C}_{\mathrm{DBC,FB}}(\varepsilon|W_1,W_2)=\mathcal{C}_{\mathrm{DBC}}(\varepsilon|W_1,W_2)\\
&=\mathcal{C}_{\mathrm{DBC}}(W_1,W_2)=\mathcal{C}(W_1,W_2).
\end{aligned}
$$

Outline of the proof of Theorem 4 will be given in the next section. The exponent function at rates outside the channel capacity in the case without feedback was derived by Arimoto [9] and Dueck and Körner [10]. The exponent function at rates outside the channel capacity in the case with feedback was derived by Csiszár and Körner [11]. They show that feedback can not improve the reliability function for the DMC at rates above capacity. The techniques used by them are not sufficient to prove Theorem 3. Some novel techniques based on the information spectrum method introduced by Han [12] are necessary to prove this theorem.

II. Outline of the Proof of the Main Result

In this section we outline the proof of Theorem 4. We first prove the following lemma.

Lemma 1: For any $\eta>0$ and for any $(\tilde{\varphi}^n,\psi_1^{(n)},\psi_2^{(n)})$ satisfying $(1/n)\log|\mathcal{K}_n|\geq R_1$, $(1/n)\log|\mathcal{L}_n|\geq R_2$, we have
$$
\begin{aligned}
&\mathrm{P}_{\mathrm{c,FB}}^{(n)}(\tilde{\varphi}^n,\psi_1^{(n)},\psi_2^{(n)})\leq\tilde{p}_{L_nX^nY^nZ^n}\Bigg\{\\
&R_1\leq\frac{1}{n}\log\frac{W_1^n(Y^n|X^n)W_2^n(Z^n|Y^n)}{q_{Y^nZ^n|L_n}(Y^n,Z^n|L_n)}+\eta, &(6)\\
&R_2\leq\frac{1}{n}\log\frac{\tilde{p}_{Z^n|L_n}(Z^n|L_n)}{\tilde{q}_{Z^n}(Z^n)}+\eta\Bigg\}+2\mathrm{e}^{-n\eta}. &(7)
\end{aligned}
$$

In (6), we can choose any conditional distribution $q_{Y^nZ^n|L_n}$ on $\mathcal{Y}^n\times\mathcal{Z}^n$ given $L_n\in\mathcal{L}_n$. In (7) we can choose any probability distribution $\tilde{q}_{Z^n}$ on $\mathcal{Z}^n$.

Proof of this lemma is given in Appendix B. For $t=1,2,\cdots,n$, set

$$\mathcal{U}_t \triangleq \mathcal{L}_n \times \mathcal{Y}^{t-1} \times \mathcal{Z}^{t-1}, \mathcal{V}_t \triangleq \mathcal{L}_n \times \mathcal{Z}^{t-1},$$
$$U_t \triangleq (L_n, Y^{t-1}, Z^{t-1}) \in \mathcal{U}_t, V_t \triangleq (L_n, Z^{t-1}) \in \mathcal{V}_t,$$
$$u_t \triangleq (l, y^{t-1}, z^{t-1}) \in \mathcal{U}_t, v_t \triangleq (l, z^{t-1}) \in \mathcal{V}_t.$$

For each $t=1,2\cdots,l$, let κ_t be a natural projection from $\mathcal{U}_t$ onto $\mathcal{V}_t$. Using κ_t, we have $V_t=\kappa_t(U_t)$, $t=1,2,\cdots,n$. For each $t=1,2,\cdots,n$, let $\tilde{\mathcal{Q}}(\mathcal{U}_t \times \mathcal{X} \times \mathcal{Y} \times \mathcal{Z})$ be a set of all probability distributions on

$$\mathcal{U}_t \times \mathcal{X} \times \mathcal{Y} \times \mathcal{Z} = \mathcal{L}_n \times \mathcal{X} \times \mathcal{Y}^t \times \mathcal{Z}^t.$$

For $t=1,2,\cdots,n$, we simply write $\tilde{\mathcal{Q}}_t=\tilde{\mathcal{Q}}(\mathcal{U}_t \times \mathcal{X} \times \mathcal{Y} \times \mathcal{Z})$. Similarly, for $t=1,2,\cdots,n$, we simply write $\tilde{q}_t=\tilde{q}_{U_tX_tY_tZ_t} \in \tilde{\mathcal{Q}}_t$. Set

$$\tilde{\mathcal{Q}}^n \triangleq \prod_{t=1}^n \tilde{\mathcal{Q}}_t = \prod_{t=1}^n \tilde{\mathcal{Q}}(\mathcal{U}_t \times \mathcal{X} \times \mathcal{Y} \times \mathcal{Z}),$$
$$\tilde{q}^n \triangleq \{\tilde{q}_t\}_{t=1}^n \in \tilde{\mathcal{Q}}^n.$$

From Lemma 1, we have the following lemma

Lemma 2: For any $\eta>0$ and for any $(\varphi^{(n)},\psi_1^{(n)},\psi_2^{(n)})$ satisfying

$$\frac{1}{n}\log|\mathcal{K}_n| \geq R_1, \frac{1}{n}\log|\mathcal{L}_n| \geq R_2.$$

we have

$$\begin{aligned}
&\mathrm{P}_{\mathrm{c,FB}}^{(n)}(\varphi^{(n)},\psi_1^{(n)},\psi_2^{(n)}) \leq \tilde{p}_{L_nX^nY^nZ^n}\Bigg\{\\
&R_1 \leq \frac{1}{n}\sum_{t=1}^n \log\frac{W_1(Y_t|X_t)}{q_{Y_t|L_nY^{t-1}}(Y_t|L_n,Y^{t-1},Z^{t-1})}+\eta,\\
&R_2 \leq \frac{1}{n}\sum_{t=1}^n \log\frac{\tilde{p}_{Z_t|L_nZ^{t-1}}(Z_t|L_n,Z^{t-1})}{\tilde{q}_{Z_t}(Z_t)}+\eta\Bigg\}+2\mathrm{e}^{-n\eta}.
\end{aligned}$$

Proof: In (6) in Lemma 1, we choose $q_{Z^nY^n|L_n}$

$$\begin{aligned}
&q_{Y^nZ^n|L_n}(y^n,z^n|l)\\
&=\prod_{t=1}^n\{q_{Y_t|L_nY^{t-1}Z^{t-1}}(y_t|l,y^{t-1},z^{t-1})\\
&\quad\times q_{Z_t|L_nY^tZ^{t-1}}(z_t|l,y^t,z^{t-1})\}\\
&=\prod_{t=1}^n\{q_{Y_t|L_nY^{t-1}Z^{t-1}}(y_t|l,y^{t-1},z^{t-1})W_2(z_t|y_t)\}.
\end{aligned}$$

In (7) in Lemma 1, we choose $\tilde{q}_{Z^n}$ having the form

$$\tilde{q}_{Z^n}(Z^n)=\prod_{t=1}^n\tilde{q}_{Z_t}(Z_t).$$

Then from the bound (7) in Lemma 1, we obtain

$$\begin{aligned}
&\mathrm{P}_{\mathrm{c,FB}}^{(n)}(\varphi^{(n)},\psi_1^{(n)},\psi_2^{(n)}) \leq \tilde{p}_{L_nX^nY^nZ^n}\Bigg\{\\
&R_1 \leq \frac{1}{n}\sum_{t=1}^n \log\frac{W_1(Y_t|X_t)}{q_{Y_t|L_nY^{t-1}Z^{t-1}}(Y_t|L_n,Y^{t-1},Z^{t-1})}+\eta,\\
&R_2 \leq \frac{1}{n}\sum_{t=1}^n \log\frac{\tilde{p}_{Z_t|L_nZ^{t-1}}(Z_t|L_n,Z^{t-1})}{\tilde{q}_{Z_t}(Z_t)}+\eta\Bigg\}+2\mathrm{e}^{-n\eta},
\end{aligned}$$

completing the proof. ■

From Lemma 2, we immediately obtain the following lemma.

Lemma 3: For any $\eta>0$, for any $(\tilde{\varphi}^n,\psi_1^{(n)},\psi_2^{(n)})$ satisfying

$$\frac{1}{n}\log|\mathcal{K}_n| \geq R_1, \frac{1}{n}\log|\mathcal{L}_n| \geq R_2,$$

and for any $\tilde{q}^n \in \tilde{\mathcal{Q}}^n$, we have

$$\begin{aligned}
&\mathrm{P}_{\mathrm{c,FB}}^{(n)}(\tilde{\varphi}^n,\psi_1^{(n)},\psi_2^{(n)}) \leq \tilde{p}_{L_nX^nY^nZ^n}\Bigg\{\\
&R_1 \leq \frac{1}{n}\sum_{t=1}^n \log\frac{W_1(Y_t|X_t)}{\tilde{q}_{Y_t|U_t}(Y_t|U_t)}+\eta,\\
&R_2 \leq \frac{1}{n}\sum_{t=1}^n \log\frac{\tilde{p}_{Z_t|V_t}(Z_t|V_t)}{\tilde{q}_{Z_t}(Z_t)}+\eta\Bigg\}+2\mathrm{e}^{-n\eta}, \qquad (8)
\end{aligned}$$

where for each $t=1,2,\cdots,n$, the conditional probability distribution $\tilde{q}_{Y_t|U_t}$ and the probability distribution $\tilde{q}_{Z_t}$ appearing in the first term in the right members of (8) are chosen so that they are induced by the joint distribution $\tilde{q}_t=\tilde{q}_{U_tX_tY_tZ_t} \in \tilde{\mathcal{Q}}_t$.

Here we define a quantity which serves as an exponential upper bound of (8) in Lemma 3. To describe this quantity we define some sets of probability distributions. Let $\mathcal{P}_{\mathrm{FB}}^{(n)}(W_1,W_2)$ be a set of all probability distributions $\tilde{p}_{L^nX^nY^nZ^n}$ on $\mathcal{L}_n \times \mathcal{X}^n \times \mathcal{Y}^n \times \mathcal{Z}^n$ having the form:

$$\begin{aligned}
&\tilde{p}_{L^nX^nY^nZ^n}(l,x^n,y^n,z^n)\\
&=\tilde{p}_{L^n}(l)\prod_{t=1}^n\{\tilde{p}_{X_t|L_nX^{t-1}Y^{t-1}Z^{t-1}}(x_t|l,x^{t-1},y^{t-1},z^{t-1})\\
&\qquad\times W_1(y_t|x_t)W_2(z_t|y_t)\}.
\end{aligned}$$

For simplicity of notation we use the notation $\tilde{p}^{(n)}$ for $\tilde{p}_{L_nX^nY^nZ^n} \in \mathcal{P}_{\mathrm{FB}}^{(n)}(W_1,W_2)$. We assume that $\tilde{p}_{U_tX_tY_tZ_t}=\tilde{p}_{L_nX_tY^tZ^t}$ is a marginal distribution of $\tilde{p}^{(n)}$. For $t=1,2,\cdots,n$, we simply write $\tilde{p}_t=\tilde{p}_{U_tX_tY_tZ_t}$. For $\tilde{p}^{(n)} \in \mathcal{P}_{\mathrm{FB}}^{(n)}(W_1,W_2)$ and $\tilde{q}^n \in \tilde{\mathcal{Q}}^n$, we define

$$\begin{aligned}
&\Omega_{\tilde{p}^{(n)}||\tilde{q}^n}^{(\mu,\theta)}(X^nY^nZ^n|L_n)\\
&\triangleq \log \mathrm{E}_{\tilde{p}^{(n)}}\left[\prod_{t=1}^n \frac{W_1^{\theta\mu}(Y_t|X_t)}{\tilde{q}_{Y_t|U_t}^{\theta\mu}(Y_t|U_t)}\frac{\tilde{p}_{Z_t|V_t}^{\theta}(Z_t|V_t)}{\tilde{q}_{Z_t}^{\theta}(Z_t)}\right],
\end{aligned}$$

where for each $t=1,2,\cdots,n$, the conditional probability distribution $\tilde{q}_{Y_t|U_t}$ and the probability distribution $\tilde{q}_{Z_t}$ appearing in the definition of $\Omega_{\tilde{p}^{(n)}||\tilde{q}^n}^{(\mu,\theta)}(X^nY^nZ^n|L_n)$ are chosen so that they are induced by the joint distribution $\tilde{q}_t=\tilde{q}_{U_tX_tY_tZ_t} \in \tilde{\mathcal{Q}}_t$. Set

$$\begin{aligned}
&\overline{\Omega}_{\mathrm{FB}}^{(\mu,\theta)}(W_1,W_2)\\
&\triangleq \sup_{n\geq 1}\max_{\tilde{p}^{(n)}\in\mathcal{P}_{\mathrm{FB}}^{(n)}(W_1,W_2)}\min_{\tilde{q}^n\in\tilde{\mathcal{Q}}^n}\frac{1}{n}\Omega_{\tilde{p}^{(n)}||\tilde{q}^n}^{(\mu,\theta)}(X^nY^nZ^n|L_n).
\end{aligned}$$

Then we have the following proposition.

Proposition 1: For any $\theta>0,\mu>0$, we have

$$G_{\mathrm{FB}}(R_1,R_2|W_1,W_2) \geq \frac{\theta(\mu R_1+R_2)-\overline{\Omega}_{\mathrm{FB}}^{(\mu,\theta)}(W_1,W_2)}{1+\theta(1+\mu)}.$$

Proof of this proposition is in Appendix C. We shall call $\overline{\Omega}_{\rm FB}^{(\mu,\theta)}(W_1,W_2)$ the communication potential. The above corollary implies that the analysis of $\overline{\Omega}_{\rm FB}^{(\mu,\theta)}(W_1,W_2)$ leads to an establishment of a strong converse theorem for the degraded BC with feedback.

The following proposition is a mathematical core to prove our main result.

Proposition 2: For $\theta \in (0,1)$, set

$$\lambda = \frac{\theta}{1-\theta} \Leftrightarrow \theta = \frac{\lambda}{1+\lambda}. \tag{9}$$

Then, for any $\theta \in (0,1)$, we have

$$\overline{\Omega}_{\rm FB}^{(\mu,\theta)}(W_1,W_2) \le \frac{1}{1+\lambda}\Omega^{(\mu,\lambda)}(W_1,W_2).$$

Proof of this proposition is in Appendix D. The proof is not so simple. We must introduce a new method for the proof.

Proof of Theorem 4: For $\theta \in (0,1)$, set

$$\lambda = \frac{\theta}{1-\theta} \Leftrightarrow \theta = \frac{\lambda}{1+\lambda}. \tag{10}$$

Then we have the following:

$$\begin{aligned} & G_{\rm FB}(R_1,R_2|W_1,W_2) \\ & \overset{\rm (a)}{\ge} \frac{\theta(\mu R_1+R_2)-\overline{\Omega}_{\rm FB}^{(\mu,\theta)}(W_1,W_2)}{1+\theta(1+\mu)} \\ & \overset{\rm (b)}{\ge} \frac{\frac{\lambda}{1+\lambda}(\mu R_1+R_2)-\frac{1}{1+\lambda}\Omega^{(\mu,\lambda)}(W_1,W_2)}{1+\frac{\lambda}{1+\lambda}(1+\mu)} \\ & = \frac{\lambda(\mu R_1+R_2)-\Omega^{(\mu,\lambda)}(W_1,W_2)}{1+\lambda+\lambda(1+\mu)} \\ & = F^{(\mu,\lambda)}(\mu R_1+R_2|W_1,W_2). \end{aligned} \tag{11}$$

Step (a) follows from Proposition 1. Step (b) follows from Proposition 2 and (10). Since (11) holds for any positive λ and μ, we have

$$G_{\rm FB}(R_1,R_2|W_1,W_2) \ge F(R_1,R_2|W_1,W_2).$$

Thus (5) in Theorem 4 is proved. ■

REFERENCES

[1] T. M. Cover, "Broadcast channels," *IEEE Trans. Inform. Theory*, vol. IT-18, no.1, pp. 2–13, Jan. 1972.

[2] P. P. Bergmans, "Random coding theorems for broadcast channels with degraded components," *IEEE Trans. Inform. Theory*, vol. IT-19, pp. 197-207, Mar. 1973.

[3] R. G. Gallager, "Capacity and coding for degraded broadcast channels," *Problemy Peredachi Informatsii*, vol. 10, pp. 3-14, July-Sept. 1974.

[4] R. F. Ahswede and J. Körner, "Source coding with side information and a converse for degraded broadcast channels," *IEEE Trans. Inform. Theory*, vol. IT-21, pp. 629-637, Nov. 1975.

[5] R. Ahlswede, P. Gàs, and J. Körner, "Bounds on conditional probabilities with applications in multi-user communication," *Z. Wahrscheinlichkeit-stheorie verw. Gebiete*, vol. 34, pp. 157-177, 1976.

[6] Y. Oohama, "Strong converse exponent for degraded broadcast channels at rates outside the capacity region," submitted for presentation at 2015 IEEE Int. Symp. on Information Theory(ISIT2015), Hong Kong, June 14-19, 2015.

[7] A. B. El Gamal, "The feedback capacity of degraded broadcast channels," *IEEE Trans. Inform. Theory*, vol. IT-24, no.3, pp.379-381, May 1978.

[8] O. Shayevitz and M. Wigger, "On the Capacity of the discrete memoryless broadcast channel with feedback, *IEEE Trans. Inform. Theory*, vol. 59, no. 3, March 2013, 1329-1345.

[9] S. Arimoto, "On the converse to the coding theorem for discrete memoryless channels," *IEEE Trans. Inform. Theory*, vol. IT-19, no. 3, pp. 357-359, May 1973.

[10] G. Dueck and J. Körner, "Reliability function of a discrete memoryless channel at rates above capacity," *IEEE Trans. Inform. Theory*, vol. IT-25, no. 1, pp. 82-85, Jan. 1979.

[11] I. Csiszár and J. Körner, "Feedback does not affect the reliability function of a DMC at rates above capacity," *IEEE Trans. Inform. Theory*, vol.IT-28, pp.92-93, 1982.

[12] T. S. Han, *Information-Spectrum Methods in Information Theory.* Springer-Verlag, Berlin, New York, 2002. The Japanese edition was published by Baifukan-publisher, Tokyo, 1998.

APPENDIX

A. Cardinality Bound of Auxilary Random Variables

We prove the following lemma.

Lemma 4: For each integer $n \ge 2$, we have

$$\begin{aligned} & \tilde{\Omega}_n^{(\mu,\lambda)}(W_1,W_2) \\ & \triangleq \max_{\substack{q=q_{UXYZ}:U\leftrightarrow X\leftrightarrow Y\leftrightarrow Z, \\ q_{Y|X}=W_1, q_{Z|Y}=W_2, \\ |\mathcal{U}|\le|\mathcal{L}_n||\mathcal{Y}|^{n-1}|\mathcal{Z}|^{n-1}}} \Omega_q^{(\mu,\lambda)}(XYZ|U) \\ & = \max_{\substack{q=q_{UXYZ}:U\leftrightarrow X\leftrightarrow Y\leftrightarrow Z, \\ q_{Y|X}=W_1, q_{Z|Y}=W_2, \\ |\mathcal{U}|\le|\mathcal{X}|}} \Omega_q^{(\mu,\lambda)}(XYZ|U) \\ & = \Omega^{(\mu,\lambda)}(W_1,W_2). \end{aligned}$$

Proof: We bound the cardinality $|\mathcal{U}|$ of U to show that the bound $|\mathcal{U}| \le |\mathcal{X}|$ is sufficient to describe $\tilde{\Omega}_n^{(\mu,\lambda)}(W_1,W_2)$. Observe that

$$q_X(x) = \sum_{u\in\mathcal{U}} q_U(u) q_{X|U}(x|u), \tag{12}$$

$$\Lambda_q^{(\mu,\lambda)}(XYZ|U) = \sum_{u\in\mathcal{U}} q_U(u)\zeta^{(\mu,\theta)}(q_{X|U}(\cdot|u)), \tag{13}$$

where

$$\begin{aligned} & \zeta^{(\mu,\lambda)}(q_{X|U}(\cdot|u)) \\ & \triangleq \sum_{(x,y,z)\in\mathcal{X}\times\mathcal{Y}\times\mathcal{Z}} q_{X|U}(x|u)W_1(y|x)W_2(z|y) \\ & \quad \times \exp\left\{\lambda\omega_q^{(\mu)}(x,y,z|u)\right\} \end{aligned}$$

are continuous functions of $q_{X|U}(\cdot|u)$. Then by the support lemma,

$$|\mathcal{U}| \le |\mathcal{X}| - 1 + 1 = |\mathcal{X}|$$

is sufficient to express $|\mathcal{X}|-1$ values of (12) and one value of (13). ■

B. Proof of Lemma 1

In this appendix we prove Lemma 1.

Proof of Lemma 1: For $l \in \mathcal{L}_n$, set

$$\begin{aligned}\tilde{\mathcal{A}}_1(l) &\stackrel{\triangle}{=} \{(x^n,y^n,z^n) : W_2^n(z^n|y^n)W_1^n(y^n|x^n) \\ &\qquad \geq |\mathcal{K}_n|\mathrm{e}^{-n\eta}q_{Y^nZ^n|L_n}(y^n,z^n|l)\},\\ \tilde{\mathcal{A}}_2(l) &\stackrel{\triangle}{=} \{(x^n,y^n,z^n) : \tilde{p}_{Z^n|L_n}(z^n|l) \geq |\mathcal{L}_n|\mathrm{e}^{-n\eta}\tilde{q}_{Z^n}(z^n)\},\\ \tilde{\mathcal{A}}(l) &\stackrel{\triangle}{=} \tilde{\mathcal{A}}_1(l) \cap \tilde{\mathcal{A}}_2(l).\end{aligned}$$

Then we have the following:

$$\begin{aligned}\mathrm{P}_{\mathrm{c,FB}}^{(n)} &= \frac{1}{|\mathcal{K}_n||\mathcal{L}_n|}\sum_{(k,l)\in\mathcal{K}_n\times\mathcal{L}_n}\sum_{\substack{(x^n,y^n,z^n)\in\tilde{\mathcal{A}}(l),\\ y^n\in\mathcal{D}_1(k),z^n\in\mathcal{D}_2(l)}} 1\\ &\quad\times\tilde{p}_{X^nY^nZ^n|K_n,L_n}(x^n,y^n,z^n|k,l)\\ &\quad+\frac{1}{|\mathcal{K}_n||\mathcal{L}_n|}\sum_{(k,l)\in\mathcal{K}_n\times\mathcal{L}_n}\sum_{\substack{(x^n,y^n,z^n)\in\tilde{\mathcal{A}}^c(l):\\ y^n\in\mathcal{D}_1(k),z^n\in\mathcal{D}_2(l)}} 1\\ &\quad\times\tilde{p}_{X^nY^nZ^n|K_n,L_n}(x^n,y^n,z^n|k,l)\\ &\leq \sum_{i=0,1,2}\tilde{\Delta}_i,\end{aligned}$$

where

$$\begin{aligned}\tilde{\Delta}_0 &\stackrel{\triangle}{=} \frac{1}{|\mathcal{K}_n||\mathcal{L}_n|}\sum_{(k,l)\in\mathcal{K}_n\times\mathcal{L}_n}\sum_{(x^n,y^n,z^n)\in\tilde{\mathcal{A}}(l)} 1\\ &\quad\times\tilde{p}_{X^nY^nZ^n|K_n,L_n}(x^n,y^n,z^n|k,l),\\ \tilde{\Delta}_i &\stackrel{\triangle}{=} \frac{1}{|\mathcal{K}_n||\mathcal{L}_n|}\sum_{(k,l)\in\mathcal{K}_n\times\mathcal{L}_n}\sum_{\substack{(x^n,y^n,z^n)\in\tilde{\mathcal{A}}_i^c(l),\\ y^n\in\mathcal{D}_1(k),z^n\in\mathcal{D}_2(l)}} 1\\ &\quad\times\tilde{p}_{X^nY^nZ^n|K_n,L_n}(x^n,y^n,z^n|k,l)\\ &\quad\text{for } i=1,2.\end{aligned}$$

By definition we have

$$\begin{aligned}&\tilde{\Delta}_0\\ &= \tilde{p}_{L_nX^nY^nZ^n}\Big\{\\ &\quad\frac{1}{n}\log|\mathcal{K}_n| \leq \frac{1}{n}\log\frac{W_1^n(Y^n|X^n)W_2^n(Z^n|Y^n)}{q_{Y^nZ^n|L_n}(Y^n,Z^n|L_n)}+\eta,\\ &\quad\frac{1}{n}\log|\mathcal{L}_n| \leq \frac{1}{n}\log\frac{\tilde{p}_{Z^n|L_n}(Z^n|L_n)}{\tilde{q}_{Z^n}(Z^n)}+\eta\Big\}.\end{aligned} \tag{14}$$

From (14), it follows that if $(\varphi^{(n)},\psi_1^{(n)},\psi_2^{(n)})$ satisfies

$$\frac{1}{n}\log|\mathcal{K}_n| \geq R_1, \frac{1}{n}\log|\mathcal{L}_n| \geq R_2,$$

then the quantity $\tilde{\Delta}_0$ is upper bounded by the first term in the right members of (7) in Lemma 1. Hence it suffices to show $\tilde{\Delta}_i \leq \mathrm{e}^{-n\eta}, i=1,2$ to prove Lemma 1. We first prove $\tilde{\Delta}_1 \leq \mathrm{e}^{-n\eta}$. We have the following chain of inequalities:

$$\begin{aligned}\tilde{\Delta}_1 &= \frac{1}{|\mathcal{K}_n||\mathcal{L}_n|}\sum_{(k,l)\in\mathcal{K}_n\times\mathcal{L}_n}\sum_{\substack{(x^n,y^n,z^n):\\ y^n\in\mathcal{D}_1(k),z^n\in\mathcal{D}_2(l)\\ W_1^n(y^n|x^n)W_2(z^n|y^n)\\ <\mathrm{e}^{-n\eta}|\mathcal{K}_n|\\ \times q_{Y^nZ^n|L_n}(y^n,z^n|l)}} 1\\ &\quad\times\tilde{\varphi}^n(x^n|k,l,y^{n-1},z^{n-1})W_1^n(y^n|x^n)W_2^n(z^n|y^n)\\ &\leq \frac{\mathrm{e}^{-n\eta}}{|\mathcal{L}_n|}\sum_{(k,l)\in\mathcal{K}_n\times\mathcal{L}_n}\sum_{\substack{(x^n,y^n,z^n):\\ y^n\in\mathcal{D}_1(k),z^n\in\mathcal{D}_2(l)}} 1\\ &\quad\times\tilde{\varphi}^n(x^n|k,l,y^{n-1},z^{n-1})q_{Y^nZ^n|L_n}(y^n,z^n|l)\\ &= \frac{\mathrm{e}^{-n\eta}}{|\mathcal{L}_n|}\sum_{(k,l)\in\mathcal{K}_n\times\mathcal{L}_n} q_{Y^nZ^n|L_n}\left(\mathcal{D}_1(k)\times\mathcal{D}_2(l)|\,l\right)\\ &\leq \frac{\mathrm{e}^{-n\eta}}{|\mathcal{L}_n|}\sum_{l\in\mathcal{L}_n}\sum_{k\in\mathcal{K}_n} q_{Y^n|L_n}\left(\mathcal{D}_1(k)|\,l\right)\\ &= \frac{\mathrm{e}^{-n\eta}}{|\mathcal{L}_n|}\sum_{l\in\mathcal{L}_n} q_{Y^n|L_n}\left(\bigcup_{k\in\mathcal{K}_n}\mathcal{D}_1(k)\middle|\,l\right)\\ &\leq \frac{\mathrm{e}^{-n\eta}}{|\mathcal{L}_n|}\sum_{l\in\mathcal{L}_n} 1 = \mathrm{e}^{-n\eta}.\end{aligned}$$

Next we prove $\tilde{\Delta}_2 \leq \mathrm{e}^{-n\eta}$. We have the following chain of inequalities:

$$\begin{aligned}\tilde{\Delta}_2 &= \frac{1}{|\mathcal{L}_n|}\sum_{(k,l)\in\mathcal{K}_n\times\mathcal{L}_n}\sum_{\substack{(x^n,y^n,z^n):\\ y^n\in\mathcal{D}_1(k),z^n\in\mathcal{D}_2(l)\\ \tilde{p}_{Z^n|L_n}<\mathrm{e}^{-n\eta}\\ \times|\mathcal{L}_n|\tilde{q}_{Z^n}(z^n)}} 1\\ &\quad\times\tilde{p}_{K_nX^nY^nZ^n|L_n}(k,x^n,y^n,z^n|l)\\ &\leq \frac{1}{|\mathcal{L}_n|}\sum_{l\in\mathcal{L}_n}\sum_{\substack{z^n\in\mathcal{D}_2(l),\\ \tilde{p}_{Z^n|L_n}<\mathrm{e}^{-n\eta}\\ \times|\mathcal{L}_n|\tilde{q}_{Z^n}(z^n)}}\sum_{k\in\mathcal{K}_n}\sum_{(x^n,y^n)\in\mathcal{X}^n\times\mathcal{Y}^n} 1\\ &\quad\times\tilde{p}_{K_nX^nY^nZ^n|L_n}(k,x^n,y^n,z^n|l)\\ &\leq \frac{1}{|\mathcal{L}_n|}\sum_{l\in\mathcal{L}_n}\sum_{\substack{z^n\in\mathcal{D}_2(l),\\ \tilde{p}_{Z^n|L_n}(z^n|l)<\mathrm{e}^{-n\eta}\\ \times|\mathcal{L}_n|\tilde{q}_{Z^n}(z^n)}} \tilde{p}_{Z^n|L_n}(z^n|l)\\ &\leq \mathrm{e}^{-n\eta}\sum_{l\in\mathcal{L}_n}\sum_{z^n\in\mathcal{D}_2(l)}\tilde{q}_{Z^n}(z^n)\\ &= \mathrm{e}^{-n\eta}\sum_{l\in\mathcal{L}_n}\tilde{q}_{Z^n}\left(\mathcal{D}_2(l)\right)\\ &= \mathrm{e}^{-n\eta}\tilde{q}_{Z^n}\left(\bigcup_{l\in\mathcal{L}_n}\mathcal{D}_2(l)\right) \leq \mathrm{e}^{-n\eta}.\end{aligned}$$

Thus Lemma 1 is proved ■

C. Proof of Proposition 1

In this appendix we prove Proposition 1. We use the following lemma, which is well known as the Cramèr's bound in the large deviation principle.

Lemma 5: For any real valued random variable Z and any $\theta > 0$, we have

$$\Pr\{Z \geq a\} \leq \exp\left[-\left(\lambda a - \log \mathrm{E}[\exp(\theta Z)]\right)\right].$$

By Lemmas 3 and 5, we have the following proposition.

Proposition 3: For any $\mu,\ \theta > 0$, any $(\tilde{\varphi}^n, \psi_1^{(n)}, \psi_2^{(n)})$ satisfying

$$\frac{1}{n}\log|\mathcal{K}_n| \geq R_1, \frac{1}{n}\log|\mathcal{L}_n| \geq R_2, \tag{15}$$

and any $\tilde{q}^n \in \tilde{\mathcal{Q}}^n$, we have

$$\begin{aligned}&\mathrm{P}_{\mathrm{c,FB}}^{(n)}(\tilde{\varphi}^n, \psi_1^{(n)}, \psi_2^{(n)})\\ &\leq 3\exp\left\{-n\frac{\theta(\mu R_1 + R_2) - \frac{1}{n}\Omega^{(\mu,\theta)}_{\tilde{p}^{(n)}||\tilde{q}^n}(X^nY^nZ^n|L_n)}{1+\theta(1+\mu)}\right\}.\end{aligned}$$

Proof: Under the condition (15), we have the following chain of inequalities:

$$\begin{aligned}&\mathrm{P}_{\mathrm{c,FB}}^{(n)}(\tilde{\varphi}^n, \psi_1^{(n)}, \psi_2^{(n)}) \overset{\text{(a)}}{\leq} \tilde{p}_{L_nX^nY^nZ^n}\Bigg\{\\ &R_1 \leq \frac{1}{n}\sum_{t=1}^{n}\log\frac{W_1(Y_t|X_t)}{\tilde{q}_{Y_t|U_t}(Y_t|U_t)} + \eta,\\ &R_2 \leq \frac{1}{n}\sum_{t=1}^{n}\log\frac{\tilde{p}_{Z_t|V_t}(Z_t|V_t)}{\tilde{q}_{Z_t}(Z_t)} + \eta\Bigg\} + 2\mathrm{e}^{-n\eta}\\ &\leq \tilde{p}_{L_nX^nY^nZ^n}\Bigg\{\mu R_1 + R_2 - (\mu+1)\eta\\ &\quad\leq \frac{1}{n}\sum_{t=1}^{n}\log\left[\frac{W_1(Y_t|X_t)\tilde{p}_{Z_t|V_t}(Z_t|V_t)}{\tilde{q}^{\mu}_{Y_t|U_t}(Y_t|U_t)\tilde{q}^{\mu}_{Z_t}(Z_t)}\right]\Bigg\} + 2\mathrm{e}^{-n\eta}\\ &\overset{\text{(b)}}{\leq} \exp\Bigg[n\Bigg\{-\theta(\mu R_1 + R_2) + \theta(\mu+1)\eta\\ &\qquad + \frac{1}{n}\Omega^{(\mu,\theta)}(X^nY^nZ^n|L_n)\Bigg\}\Bigg] + 2\mathrm{e}^{-n\eta}.\end{aligned} \tag{16}$$

Step (a) follows from Lemma 3. Step (b) follows from Lemma 5. We choose η so that

$$\begin{aligned}-\eta = &-\theta(\mu R_1 + R_2) + \theta(\mu+1)\eta\\ &+ \frac{1}{n}\Omega^{(\mu,\theta)}_{\tilde{p}^{(n)}||\tilde{q}^n}(X^nY^nZ^n|L_n).\end{aligned} \tag{17}$$

Solving (17) with respect to η, we have

$$\eta = \frac{\theta(\mu R_1 + R_2) - \frac{1}{n}\Omega^{(\mu,\theta)}_{\tilde{p}^{(n)}||\tilde{q}^n}(X^nY^nZ^n|L_n)}{1+\theta(1+\mu)}.$$

For this choice of η and (16), we have

$$\begin{aligned}&\mathrm{P}_{\mathrm{c,FB}}^{(n)} \leq 3\mathrm{e}^{-n\eta}\\ &= 3\exp\left\{-n\frac{\theta(\mu R_1 + R_2) - \frac{1}{n}\Omega^{(\mu,\theta)}_{\tilde{p}^{(n)}||\tilde{q}^n}(X^nY^nZ^n|L_n)}{1+\theta(1+\mu)}\right\},\end{aligned}$$

completing the proof. ∎

Proof of Proposition 1 By the definitions of $G^{(n)}_{\mathrm{FB}}(R_1, R_2|W_1, W_2)$ and $\overline{\Omega}^{(\mu,\theta)}_{\mathrm{FB}}(W_1, W_2)$ and Proposition 3, we have

$$\begin{aligned}&G^{(n)}_{\mathrm{FB}}(R_1, R_2|W_1, W_2)\\ &\geq \frac{\theta(\mu R_1 + R_2) - \overline{\Omega}^{(\mu,\theta)}_{\mathrm{FB}}(W_1, W_2)}{1+\theta(1+\mu)} - \frac{1}{n}\log 3.\end{aligned} \tag{18}$$

From (18), we have Proposition 1. ∎

D. Upper Bound of $\overline{\Omega}^{(\mu,\theta)}_{\mathrm{FB}}(W_1, W_2)$

In this appendix we drive an explicit upper bound of $\overline{\Omega}^{(\mu,\theta)}_{\mathrm{FB}}(W_1, W_2)$ to prove Proposition 2. For each $t = 1, 2, \cdots, n$, define the function of $(u_t, x_t, y_t, z_t) \in \mathcal{V}_t \times \mathcal{X} \times \mathcal{Y} \times \mathcal{Z}$ by

$$f^{(\mu,\lambda)}_{\tilde{p}_t||\tilde{q}_t,\kappa_t}(x_t, y_t, z_t|u_t) \triangleq \frac{W_1^{\theta\mu}(y_t|x_t)\tilde{p}^{\theta}_{Z_t|V_t}(z_t|v_t)}{\tilde{q}^{\theta\mu}_{Y_t|U_t}(y_t|u_t)\tilde{q}^{\theta}_{Z_t}(z_t)}.$$

For each $t = 1, 2, \cdots, n$, we define the probability distribution

$$\begin{aligned}&\tilde{p}^{(\mu,\theta;\tilde{q}^t,\kappa^t)}_{L_nX^tY^tZ^t}\\ &\triangleq \left\{\tilde{p}^{(\mu,\theta;\tilde{q}^t,\kappa^t)}_{L_nX^tY^tZ^t}(l, x^t, y^t, z^t)\right\}_{(l,x^t,y^t,z^t)\in\mathcal{L}_n\times\mathcal{X}^t\times\mathcal{Y}^t\times\mathcal{Z}^t}\end{aligned}$$

by

$$\begin{aligned}&\tilde{p}^{(\mu,\theta;\tilde{q}^t,\kappa^t)}_{L_nX^tY^tZ^t}(l, x^t, y^t, z^t)\\ &\triangleq \tilde{C}_t^{-1}p_{L_n}(l)\prod_{i=1}^{t}\{\tilde{p}_{X_i|U_iX^{i-1}}(x_i|u_i, x^{i-1})\\ &\quad\times W_1(y_i|x_i)W_2(z_i|y_i)f^{(\mu,\theta)}_{\tilde{p}_i||\tilde{q}_i,\kappa_i}(x_i, y_i, z_i|u_i)\},\end{aligned}$$

where

$$\begin{aligned}\tilde{C}_t \triangleq &\sum_{l,x^t,y^t,z^t}p_{L_n}(l)\prod_{i=1}^{t}\{\tilde{p}_{X_i|U_iX^{i-1}}(x_i|u_i, x^{i-1})\\ &\times W_1(y_i|x_i)W_2(z_i|y_i)f^{(\mu,\theta)}_{\tilde{p}_i||\tilde{q}_i,\kappa_i}(x_i, y_i, z_i|u_i)\}.\end{aligned}$$

are constants for normalization. For each $t = 1, 2, \cdots, n$, set

$$\tilde{\Phi}^{(\mu,\theta)}_{t,\tilde{q}^t,\kappa^t} \triangleq \tilde{C}_t\tilde{C}_{t-1}^{-1}, \tag{19}$$

where we define $\tilde{C}_0 = 1$. Then we have the following lemma.

Lemma 6:

$$\Omega^{(\mu,\theta)}_{\tilde{p}^{(n)}||\tilde{q}^n}(X^nY^nZ^n|L_n) = \sum_{t=1}^{n}\log\tilde{\Phi}^{(\mu,\theta)}_{t,\tilde{q}^t,\kappa^t}. \tag{20}$$

Proof: From (19) we have

$$\log\tilde{\Phi}^{(\mu,\theta)}_{t,\tilde{q}^t,\kappa^t} = \log\tilde{C}_t - \log\tilde{C}_{t-1}. \tag{21}$$

Furthermore, by definition we have

$$\Omega^{(\mu,\theta)}_{\tilde{p}^{(n)}||\tilde{q}^n}(X^nY^nZ^n|L_n) = \log\tilde{C}_n, \tilde{C}_0 = 1. \tag{22}$$

From (21) and (22), (20) is obvious. ∎

The following lemma is useful for the computation of $\tilde{\Phi}^{(\mu,\theta)}_{t,\tilde{q}^t,\kappa^t}$ for $t = 1, 2, \cdots, n$.

Lemma 7: For each $t = 1,2,\cdots,n$, and for any $(l, x^t, y^t, z^t) \in \mathcal{L}_n \times\mathcal{X}^t \times\mathcal{Y}^t \times\mathcal{Z}^t$, we have

$$\begin{aligned}&\tilde{p}^{(\mu,\theta;\tilde{q}^t,\kappa^t)}_{L_nX^tY^tZ^t}(l,x^t,y^t,z^t)\\&=(\tilde{\Phi}^{(\mu,\theta)}_{t,\tilde{q}^t,\kappa^t})^{-1}\tilde{p}^{(\mu,\theta;\tilde{q}^{t-1},\kappa^{t-1})}_{L_nX^{t-1}Y^{t-1}Z^{t-1}}(l,x^{t-1},y^{t-1},z^{t-1})\\&\quad\times\tilde{p}_{X_t|U_tX^{t-1}}(x_t|u_t,x^{t-1})W_1(y_t|x_t)W_2(z_t|y_t)\\&\quad\times f^{(\mu,\theta)}_{\tilde{p}_t||\tilde{q}_t,\kappa_t}(x_t,y_t,z_t|u_t). \end{aligned}\tag{23}$$

Furthermore, we have

$$\begin{aligned}\tilde{\Phi}^{(\mu,\theta)}_{t,\tilde{q}^t,\kappa^t} &= \sum_{l,x^t,y^t,z^t}\tilde{p}^{(\mu,\theta;\tilde{q}^{t-1},\kappa^{t-1})}_{L_nX^{t-1}Y^{t-1}Z^{t-1}}(l,x^{t-1},y^{t-1},z^{t-1})\\&\quad\times\tilde{p}_{X_t|U_tX^{t-1}}(x_t|u_t,x^{t-1})W_1(y_t|x_t)W_2(z_t|y_t)\\&\quad\times f^{(\mu,\theta)}_{\tilde{p}_t||\tilde{q}_t,\kappa_t}(x_t,y_t,z_t|u_t)\\&=\sum_{u_t,x^t,y_t,z_t}\tilde{p}^{(\mu,\theta;\tilde{q}^{t-1},\kappa^{t-1})}_{U_tX^{t-1}}(u_t,x^{t-1})\\&\quad\times\tilde{p}_{X_t|U_tX^{t-1}}(x_t|u_t,x^{t-1})W_1(y_t|x_t)W_2(z_t|y_t)\\&\quad\times f^{(\mu,\theta)}_{\tilde{p}_t||\tilde{q}_t,\kappa_t}(x_t,y_t,z_t|u_t).\end{aligned}\tag{24}$$

Proof: By the definition of $\tilde{p}^{(\mu,\theta;\tilde{q}^t,\kappa^t)}_{L_nX^tY^tZ^t}$ (l, x^t, y^t, z^t), $t = 1,2,\cdots,n$, we have

$$\begin{aligned}&\tilde{p}^{(\mu,\theta;\tilde{q}^t,\kappa^t)}_{L_nX^tY^tZ^t}(l,x^t,y^t,z^t)\\&=\tilde{C}_t^{-1}p_{L_n}(l)\prod_{i=1}^{t}\{\tilde{p}_{X_i|U_iX^{i-1}}(x_i|u_i,x^{i-1})\\&\quad\times W_1(y_i|x_i)W_2(z_i|y_i)f^{(\mu,\theta)}_{\tilde{p}_i||\tilde{q}_i,\kappa_i}(x_i,y_i,z_i|u_i)\}.\end{aligned}\tag{25}$$

Then we have the following chain of equalities:

$$\begin{aligned}&\tilde{p}^{(\mu,\theta;\tilde{q}^t,\kappa^t)}_{L_nX^tY^tZ^t}(l,x^t,y^t,z^t)\\&\overset{\text{(a)}}{=}\tilde{C}_t^{-1}p_{L_n}(l)\prod_{i=1}^{t}\{\tilde{p}_{X_i|U_iX^{i-1}}(x_i|u_i,x^{i-1})\\&\quad\times W_1(y_i|x_i)W_2(z_i|y_i)f^{(\mu,\theta)}_{\tilde{p}_i||\tilde{q}_i,\kappa_i}(x_i,y_i,z_i|u_i)\}\\&=\tilde{C}_t^{-1}p_{L_n}(l)\prod_{i=1}^{t-1}\{\tilde{p}_{X_i|U_iX^{i-1}}(x_i|u_i,x^{i-1})\\&\quad\times W_1(y_i|x_i)W_2(z_i|y_i)f^{(\mu,\theta)}_{\tilde{p}_i||\tilde{q}_i,\kappa_i}(x_i,y_i,z_i|u_i)\}\\&\quad\times\tilde{p}_{X_t|U_tX^{t-1}}(x_t|u_t,x^{t-1})W_1(y_t|x_t)W_2(z_t|y_t)\\&\quad\times f^{(\mu,\theta)}_{\tilde{p}_t||\tilde{q}_t,\kappa_t}(x_t,y_t,z_t|u_t)\\&\overset{\text{(b)}}{=}\tilde{C}_t^{-1}\tilde{C}_{t-1}\tilde{p}^{(\mu,\theta;\tilde{q}^{t-1},\kappa^{t-1})}_{L_nX^{t-1}Y^{t-1}Z^{t-1}}(l,x^{t-1},y^{t-1},z^{t-1})\\&\quad\times\tilde{p}_{X_t|U_tX^{t-1}}(x_t|u_t,x^{t-1})W_1(y_t|x_t)W_2(z_t|y_t)\\&\quad\times f^{(\mu,\theta)}_{\tilde{p}_t||\tilde{q}_t,\kappa_t}(x_t,y_t,z_t|u_t)\\&=(\tilde{\Phi}^{(\mu,\theta)}_{t,\tilde{q}^t,\kappa^t})^{-1}\tilde{p}^{(\mu,\theta;\tilde{q}^{t-1},\kappa^{t-1})}_{L_nX^{t-1}Y^{t-1}Z^{t-1}}(l,x^{t-1},y^{t-1},z^{t-1})\\&\quad\times\tilde{p}_{X_t|U_tX^{t-1}}(x_t|u_t,x^{t-1})W_1(y_t|x_t)W_2(z_t|y_t)\\&\quad\times f^{(\mu,\theta)}_{\tilde{p}_t||\tilde{q}_t,\kappa_t}(x_t,y_t,z_t|u_t).\end{aligned}\tag{26}$$

Steps (a) and (b) follow from (25). From (26), we have

$$\tilde{\Phi}^{(\mu,\theta)}_{t,\tilde{q}^t,\kappa^t}\tilde{p}^{(\mu,\theta;\tilde{q}^t,\kappa^t)}_{L_nX^tY^tZ^t}(l,x^t,y^t,z^t)\tag{27}$$

$$\begin{aligned}&=\tilde{p}^{(\mu,\theta;\tilde{q}^{t-1},\kappa^{t-1})}_{L_nX^{t-1}Y^{t-1}Z^{t-1}}(l,x^{t-1},y^{t-1},z^{t-1})\\&\quad\times\tilde{p}_{X_t|U_tX^{t-1}}(x_t|u_t,x^{t-1})W_1(y_t|x_t)W_2(z_t|y_t)\\&\quad\times f^{(\mu,\theta)}_{\tilde{p}_t||\tilde{q}_t,\kappa_t}(x_t,y_t,z_t|u_t).\end{aligned}\tag{28}$$

Taking summations of (27) and (28) with respect to l,x^t, y^t, z^t, we obtain

$$\begin{aligned}&\tilde{\Phi}^{(\mu,\theta)}_{t,\tilde{q}^t,\kappa^t}\\&=\sum_{l,x^t,y^t,z^t}\tilde{p}^{(\mu,\theta;\tilde{q}^{t-1},\kappa^{t-1})}_{L_nX^{t-1}Y^{t-1}Z^{t-1}}(l,x^{t-1},y^{t-1},z^{t-1})\\&\quad\times\tilde{p}_{X_t|U_t,X^{t-1}}(x_t|u_t,x^{t-1})W_1(y_t|x_t)W_2(z_t|y_t)\\&\quad\times f^{(\mu,\theta)}_{\tilde{p}_t||\tilde{q}_t,\kappa_t}(x_t,y_t,z_t|u_t)\\&=\sum_{u_t,x^t,y_t,z_t}\tilde{p}^{(\mu,\theta;\tilde{q}^{t-1},\kappa^{t-1})}_{U_tX^{t-1}}(u_t,x^{t-1})\\&\quad\times\tilde{p}_{X_t|U_t,X^{t-1}}(x_t|u_t,x^{t-1})W_1(y_t|x_t)W_2(z_t|y_t)\\&\quad\times f^{(\mu,\theta)}_{\tilde{p}_t||\tilde{q}_t,\kappa_t}(x_t,y_t,z_t|u_t),\end{aligned}$$

completing the proof. ∎

We set

$$\begin{aligned}&\tilde{p}^{(\mu,\theta;\tilde{q}^{t-1},\kappa^{t-1})}_{U_tX_t}(u_t,x_t)\\&=\sum_{x^{t-1}}\tilde{p}^{(\mu,\theta;\tilde{q}^{t-1},\kappa^{t-1})}_{U_tX^{t-1}}(u_t,x^{t-1})\tilde{p}_{X_t|U_tX^{t-1}}(x_t|u_t,x^{t-1}).\end{aligned}$$

Then by (24) in Lemma 7 and the definition of $f^{(\mu,\theta)}_{\tilde{p}_t||\tilde{q}_t,\kappa_t}$ $(x_t,y_t,z_t|u_t)$, we have

$$\begin{aligned}&\tilde{\Phi}^{(\mu,\theta)}_{t,\tilde{q}^t,\kappa^t}\\&=\sum_{u_t,x_t,y_t,z_t}\tilde{p}^{(\mu,\theta;\tilde{q}^{t-1},\kappa^{t-1})}_{U_tX_t}(u_t,x_t)W_1(y_t|x_t)W_2(z_t|y_t)\\&\quad\times\frac{W_1^{\theta\mu}(y_t|x_t)\tilde{p}^{\theta}_{Z_t|V_t}(z_t|v_t)}{\tilde{q}^{\theta\mu}_{Y_t|U_t}(y_t|u_t)\tilde{q}^{\theta}_{Z_t}(z_t)}.\end{aligned}\tag{29}$$

Proof of Proposition 2 is as follows.

Proof of Proposition 2: Set

$$\begin{aligned}\tilde{\mathcal{P}}_n(W_1,W_2)&\stackrel{\triangle}{=}\{\tilde{q}=\tilde{q}_{UXYZ}:|\mathcal{U}|\leq|\mathcal{L}_n||\mathcal{Y}|^{n-1}|\mathcal{Z}|^{n-1},\\&\qquad\tilde{q}_{Y|X}=W_1,\tilde{q}_{Z|Y}=W_2,U\leftrightarrow X\leftrightarrow Y\leftrightarrow Z\},\\\tilde{\Omega}_n^{(\mu,\lambda)}(W_1,W_2)&\stackrel{\triangle}{=}\max_{\tilde{q}\in\tilde{\mathcal{P}}_n(W_1,W_2)}\log\Omega^{(\mu,\lambda)}_{\tilde{q}}(XYZ|U).\end{aligned}$$

We choose $\tilde{q}_t = \tilde{q}_{U_tX_tY_tZ_t}$ so that

$$\begin{aligned}&\tilde{q}_{U_tX_tY_tZ_t}(u_t,x_t,y_t,z_t)\\&=\tilde{p}^{(\mu,\theta;\tilde{q}^{t-1},\kappa^{t-1})}_{U_tX_t}(u_t,x_t)W_1(y_1|x_t)W_2(z_t|y_t).\end{aligned}$$

It is obvious that $\tilde{q}_t \in \tilde{\mathcal{P}}_n(W_1, W_2)$ for $t = 1, 2, \cdots, n$. By (29) and the above choice of $\tilde{q}_t$, we have

$$
\begin{aligned}
&\tilde{\Phi}^{(\mu,\theta)}_{t,\tilde{q}^t,\kappa^t} \\
&= \sum_{u_t,x^t,y_t,z_t} \tilde{q}_{U_t}(u_t)\tilde{q}_{X_t|U_t}(x_t|u_t)W_1(y_t|x_t)W_2(z_t|y_t). \\
&\quad \times \left\{ \frac{W_1^{\mu}(y_t|x_t)}{\tilde{q}^{\mu}_{Y_t|U_t}(y_t|u_t)} \frac{\tilde{p}_{Z_t|V_t}(z_t|v_t)}{\tilde{q}_{Z_t}(z_t)} \right\}^{\theta} \\
&= \mathrm{E}_{\tilde{q}_t}\left[\left\{ \frac{W_1^{\mu}(Y_t|X_t)}{\tilde{q}^{\mu}_{Y_t|U_t}(Y_t|U_t)} \frac{\tilde{p}_{Z_t|V_t}(Z_t|V_t)}{\tilde{q}_{Z_t}(V_t)} \right\}^{\theta} \right] \\
&= \mathrm{E}_{\tilde{q}_t}\left[\left\{ \frac{W_1^{\mu}(Y_t|X_t)}{\tilde{q}^{\mu}_{Y_t|U_t}(Y_t|U_t)} \frac{\tilde{q}_{Z_t|U_t}(Z_t|U_t)}{\tilde{q}_{Z_t}(Z_t)} \frac{\tilde{p}_{Z_t|V_t}(Z_t|V_t)}{\tilde{q}_{Z_t|U_t}(Z_t|U_t)} \right\}^{\theta} \right] \\
&\overset{\text{(a)}}{\leq} \left(\mathrm{E}_{\tilde{q}_t}\left[\left\{ \frac{W_1^{\mu}(Y_t|X_t)}{\tilde{q}^{\mu}_{Y_t|U_t}(Y_t|U_t)} \frac{\tilde{q}_{Z_t|U_t}(Z_t|U_t)}{\tilde{q}_{Z_t}(Z_t)} \right\}^{\frac{\theta}{1-\theta}} \right] \right)^{1-\theta} \\
&\quad \times \left(\mathrm{E}_{\tilde{q}_t}\left\{ \frac{\tilde{p}_{Z_t|V_t}(Z_t|V_t)}{\tilde{q}_{Z_t|U_t}(Z_t|U_t)} \right\} \right)^{\theta} \\
&= \exp\left\{ (1-\theta)\Omega^{(\mu,\frac{\theta}{1-\theta})}_{\tilde{q}_t}(X_tY_tZ_t|U_t) \right\} \\
&\overset{\text{(b)}}{=} \exp\left\{ \frac{1}{1+\lambda}\Omega^{(\mu,\lambda)}_{\tilde{q}_t}(X_tY_tZ_t|U_t) \right\} \\
&\overset{\text{(c)}}{\leq} \exp\left\{ \frac{1}{1+\lambda}\tilde{\Omega}^{(\mu,\lambda)}_{n}(W_1,W_2) \right\} \\
&\overset{\text{(d)}}{=} \exp\left\{ \frac{1}{1+\lambda}\Omega^{(\mu,\lambda)}(W_1,W_2) \right\}. \qquad (30)
\end{aligned}
$$

Step (a) follows from Hölder's inequality. Step (b) follows from (9). Step (c) follows from $\tilde{q}_t \in \tilde{\mathcal{P}}_n(W_1, W_2)$ and the definition of $\tilde{\Omega}^{(\mu,\lambda)}_n(W_1, W_2)$. Step (d) follows from Lemma 4 in Appendix A. To prove this lemma we bound the cardinality $|\mathcal{V}|$ appearing in the definition of $\tilde{\Omega}^{(\mu,\lambda)}_n(W_1, W_2)$ to show that the bound $|\mathcal{U}| \leq |\mathcal{X}|$ is sufficient to describe $\tilde{\Omega}^{(\mu,\lambda)}_n(W_1, W_2)$. Hence we have the following:

$$
\begin{aligned}
&\min_{\tilde{q}^n \in \tilde{\mathcal{Q}}^n} \frac{1}{n}\Omega^{(\mu,\theta)}_{\tilde{p}^{(n)}||\tilde{q}^n}(X^nY^nZ^n|L_n) \\
&\leq \frac{1}{n}\Omega^{(\mu,\theta)}_{\tilde{p}^{(n)}||\tilde{q}^n}(X^nY^nZ^n|L_n) \overset{\text{(a)}}{=} \frac{1}{n}\sum_{t=1}^{n}\log \tilde{\Phi}^{(\mu,\theta)}_{t,\tilde{q}^t,\kappa^t} \\
&\overset{\text{(b)}}{\leq} \frac{1}{1+\lambda}\Omega^{(\mu,\lambda)}(W_1,W_2). \qquad (31)
\end{aligned}
$$

Step (a) follows from (20) in Lemma 6. Step (b) follows from (30). Since (31) holds for any $n \geq 1$ and any $\tilde{p}^{(n)} \in \mathcal{P}^{(n)}_{\mathrm{FB}}$ (W_1, W_2), we have

$$
\overline{\Omega}^{(\mu,\theta)}_{\mathrm{FB}}(W_1,W_2) \leq \frac{1}{1+\gamma}\Omega^{(\mu,\lambda)}(W_1,W_2).
$$

Thus, Proposition 2 is proved. ■